Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 36 (2009) 933–943
Expert Systems with Applications
www.elsevier.com/locate/eswa

A Bayesian network and analytic hierarchy process based personalized recommendations for tourist attractions over the Internet

Yuxia Huang *, Ling Bian

Department of Geography, State University of New York at Buffalo, Buffalo, NY 14260, USA

Abstract

Selecting tourist attractions to visit at a destination is a main stage in planning a trip. Although various online travel recommendation systems have been developed to support users in the task of travel planning during the last decade, few systems focus on recommending specific tourist attractions. In this paper, an intelligent system to provide personalized recommendations of tourist attractions in an unfamiliar city is presented. Through a tourism ontology, the system allows integration of heterogeneous online travel information. Based on Bayesian network technique and the analytic hierarchy process (AHP) method, the system recommends tourist attractions to a user by taking into account the travel behavior both of the user and of other users. Spatial web services technology is embedded in the system to provide GIS functions. In addition, the system provides an interactive geographic interface for displaying the recommendation results as well as obtaining users' feedback. The experiments show that the system can provide personalized recommendations on tourist attractions that satisfy the user.

Keywords: Personalized recommendation; Tourist attractions; Ontology; Bayesian network; Analytic hierarchy process

1. Introduction

Personalized recommendation is one-to-one recommendation through understanding each persona individually (Good et al., 1999; Resnick & Varian, 1997). Recently personalized recommendation systems have been gaining interest in tourism to assist travelers with travel plans (Loh, Lorenzi, Saldana, & Licthnow, 2003; Ricci, 2002; Ricci & Werthner, 2002; Wallace, Maglogiannis, Karpouzis, Kormentzas, & Kollias, 2003). A travel plan consists of a number of stages, such as choosing destinations, selecting tourist attractions, choosing accommodations, deciding routes, etc. However, at present most of travel recommendations focus on the first stage – suggesting the destinations, with very few exceptions (Ardissono, Goy, Petrone, Signan, & Torasso, 2003). This paper is concerned with the second stage, which suggests a set of tourist attractions in sequence at a given destination. Tourist attractions, which are places intended to attract people to visit at a destination, are often the reason driving travelers to visit destinations (Gunn, 1980; Leiper, 1990; Lew, 1987; Jafari, 2000; Richards, 2002).

There are two challenges in developing a system for personalized recommendations of visiting tourist attractions. One is the integration of heterogeneous online travel information. The other is the semantic matching tourist attractions against travelers' preferences. First, this recommendation process involves a large amount of detailed up-to-date information of tourist attractions (Ardissono et al., 2003; Fesenmaier & Jeng, 2000). The information is available over the Internet, published by various travel information providers. However, due to the heterogeneity of the information, it is difficult to automatically integrate the information. Different providers may use different terms to represent the same meaning, or same terms for different meanings. Furthermore, in order to recommend satisfactory tourist attractions to travelers, the characteristics of tourist attractions, for example the activities offered in an attraction, have to match the

* Corresponding author. Tel.: +1 716 645 2722x44.
E-mail address: huang3@buffalo.edu (Y. Huang).

0957-4174/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2007.10.019

travelers' preferences. However, travel preferences are often hidden and are not explicitly known when users start to plan their trips, particularly if visiting an unfamiliar place (Loh et al., 2003; Viappiani, Pu, & Faltings, 2002). Thus, in this matching process, two stages are involved: estimating travelers' preferences, and subsequently evaluating available tourist attractions.

In this paper, a system for personalized recommendations of tourist attractions in a travel destination is presented. The system has four main characteristics. First, information of tourist attractions in the destination is obtained by understanding and integrating heterogeneous online information based on a tourism ontology. The ontology contains a set of concepts to express the characteristics of tourist attractions and the relationships between these concepts that the system needs to generate recommendations of tour plans. It is concerned with the services provided at tourist attractions. Second, a Bayesian network is used to estimate the traveler's preferred activities. With the Bayesian network, travel behavior of the person and of other travelers who have similar taste can be combined. Third, the analytic hierarchy process (AHP) is used to rank the available tourist attractions because of its capability of using the traveler's travel behavior in the past. Finally, spatial web services technology is used to incorporate on-demand mapping and spatial functionality into the system.

The rest of this paper is organized as follows. Section 2 describes the ontological modeling of the information of tourist attractions. The estimation of the traveler's preferred activities using a Bayesian network and ranking the tourist attractions using AHP are discussed in Sections 3 and 4, respectively. Section 5 describes embedding GIS functions using spatial web services. The architecture of the system and its implementation are presented in Section 6. Finally, the summary of this research is discussed in Section 7.

2. Ontological data model for information about tourist attractions

2.1. Ontology for integrating heterogeneous information about tourist attractions

Ontology, which is a theory about the nature of existence in philosophy, plays an important role in the integrating heterogeneous information (Dell' Erba, Fodor, Hopken, & Werthner, 2005; Gruber, 1993; Hyvonen et al., 2005). Various definitions of ontology have been given by researchers. The most commonly quoted definition of an ontology is "*a formal, explicit specification of a shared conceptualization* "(Gruber, 1993). Here, an *explicit specification* means that the concepts and relationships in the abstract model are given explicit names and definitions, which are agreed upon cross communities. In other words, an ontology describes a shared and common understanding of a domain that can be communicated among communities.

Concepts and relationships are basic components in an ontology. Web documents are the most important source for deriving concepts and relationships (Maedche & Staab, 2001). For this research, the online information of tourist attractions is the source to derive concepts and the relationships between these concepts in the tourist ontology. By examining more than 200 tourism websites for attractions at Niagara Falls and New York City, we found that the main investigated concepts in this ontology are 'Attraction' (the basic information such its as name), 'Location' (the geographic information), 'Open Times' (the temporal information), 'Admission Fees' (the cost), and 'Activity' (things can be done while visiting an attraction). Each concept may have a set of sub concepts. For example, 'Open Times' has three sub concepts: 'Open Hours' (open schedule), 'Closed Dates' (closed date), and 'Minimum Time to Stay' (the minimum time required to stay at an attraction).

Fig. 1 illustrates these concepts and the relationships between them. A tourist attraction has an unique location and may offer many activities. An activity has open time and admission fees. Open time includes open hours, closed date, and minimum time to stay at an attraction. Each activity may have different admission fees, depending on travelers' age group and occupation (e.g., student).

2.2. Representing the ontology in OWL

For this research, Web Ontology Language (OWL),[1] recommended by the World Wide Web Consortium (W3C), is used to represent the ontology due to its capability of explicitly representing the concepts and their relationships.

In OWL, class, instance and property are the three main components. A class is used to represent a concept defined in the ontology. Classes are typically arranged in a taxonomy, including classes and subclasses. A class may associate with a set of instances. A property is used to represent the concept's attributes, including the relationships with other concepts. Each property is associated with domain, range, cardinality, minCardinality, and maxCardinality. A domain of a property limits the class to which the property can be applied, while a range of a property limits the class that the property may have as its value. Cardinality is used to describe the number of values for a class, and minCardinality and maxCardinality describe the minimum and maximum number of values for a class, respectively. Table 1 shows the example properties used to represent the relationships between concepts, and Table 2 describes the part of OWL code in representing the concept 'Attraction'. For example, the property "providedActivity" is applied to the class "Activity". Its values are from the class "Activity". An attraction must offer at least one activity, but may offer many.

[1] http://www.w3.org/.